remote sensing

MDPI

Article

The potential of Landsat NDVI sequences explain wheat yield variation in fields in Western Australia

Jianxiu Shen [1] **and Fiona H. Evans** [1,2,*]

[1] Centre for Crop and Food Innovation, Food Futures Institute, Murdoch University, 90 South Street, Murdoch, WA Australia 6150
[2] Centre for Digital Agriculture, Centre for Crop and Disease Management, Curtin University, Kent Street, Bentley, WA Australia 6102
* Correspondence: Fiona.Evans@murdoch.edu.au

Supplementary Material File S1: For training field J, 1) examples and justification for the major steps in the statistical phenology detection (SPD) and for the conditional settings in the use of sow dates (+SD) and break of season (+BOS) before performing SPD; 2) the selection process of integrated NDVI metrics for yield estimation using SPD, +SD, and +BOS.

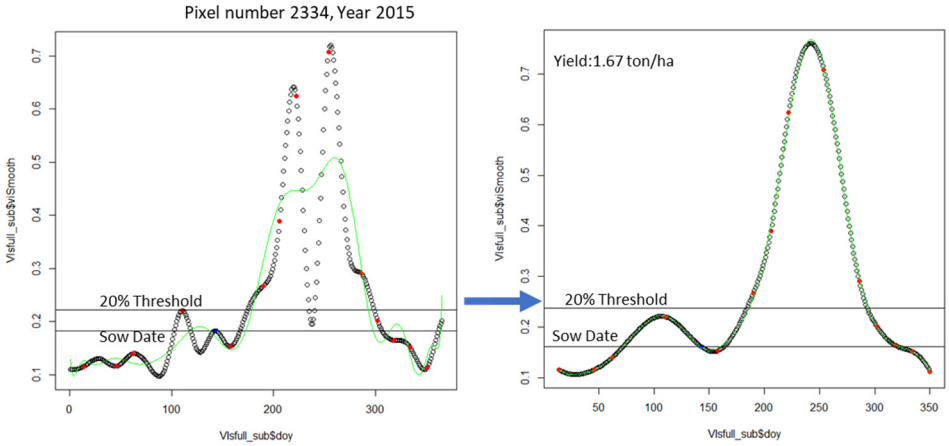


Figure S1 An example in field J of applying the first step 'interpolate and fill the missing values to create a regular 16-day time sequence of NDVI' before interpolating to daily in section 3.1.1. red dots are the original available NDVI time points from Landsat 7 imagery. Black dots are the smoothed daily NDVI points. Green lines are the fitted curve by multi-polynomial fitting function. Pixel number was ordered by rows then columns of the gridded field. The left panel is the daily reconstructed curve without implementing the first step. It shows unreasonable curve estimation in the middle of the growing season. The right panel demonstrates the daily reconstruction results from applying 16-day reconstruction step beforehand.

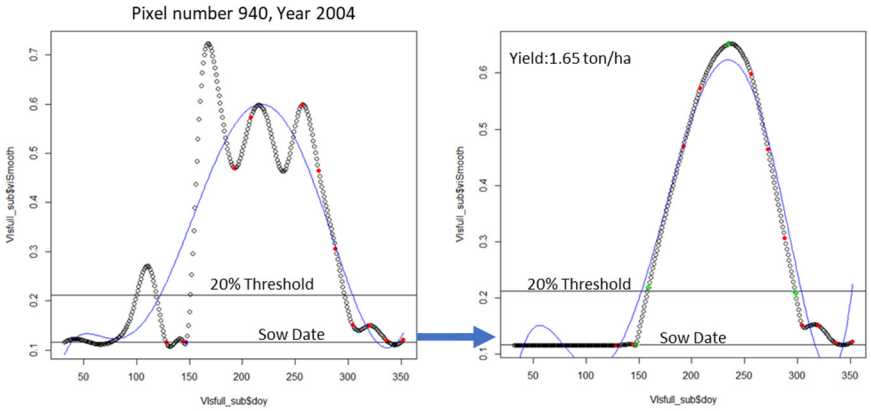


Figure S2 An example of using the Stineman interpolation method in step 1 and 2 in section 3.1.1. The left and right panels show the daily reconstructed results using 'spline'[1] and 'stineman' [2] functions, respectively. Both functions have advantages and limitations in different circumstances. The temporal distribution of data for this pixel in year 2004 shows less available points in the early growing season but relatively more points in the late growing season. 'stineman' interpolation function showed better performance for this pattern of data.

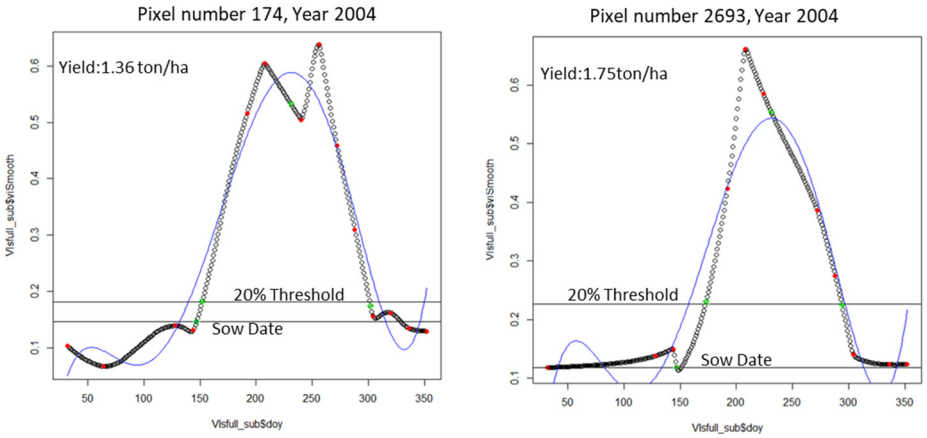


Figure S3 Two example pixels for showing the identification of peak NDVI date based on the fitted multi-polynomial curve (step 3 in section 3.1.1). There was a sharp drop of NDVI in the middle of the growing season in year 2004. The drop was occurred at most of the pixels and at field average level (Figure S7). The fitting function not only helps to define the stages of continuous NDVI increase and decrease (Step 4 in section 3.1.1), but also helps to locate the date when peak of season NDVI occurs.

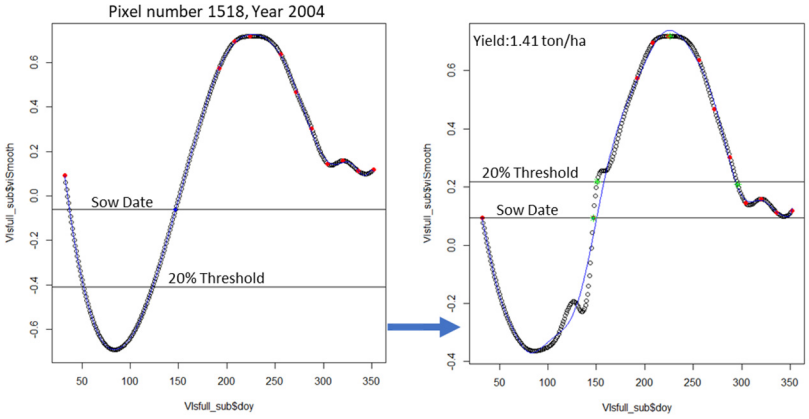


Figure S4 An example of using the original time sequence of data points to calculate the 20% threshold and sow date NDVI, rather than using the smoothed sequence.

As shown in the left panel of Figure S4, the smoothed time sequence has non-realistic NDVI values that are lower than 0. This was caused by the lack of available data points in prior to and during the first half of the growing season. Consequently, the values of 20% threshold and sowing date NDVI were lower than 0 if we use the smoothed time sequence. The right panel shows the corrected curve after 1) adding the sowing date NDVI based on the minimum value of the original available sequence (step 2 in section 3.1.2); 2) calculating the 20% threshold as 20% of the original sequence amplitude (step 4 in section 3.1.1). After these corrections, the phenological metrics that describe the growing season were reasonable and acceptable.

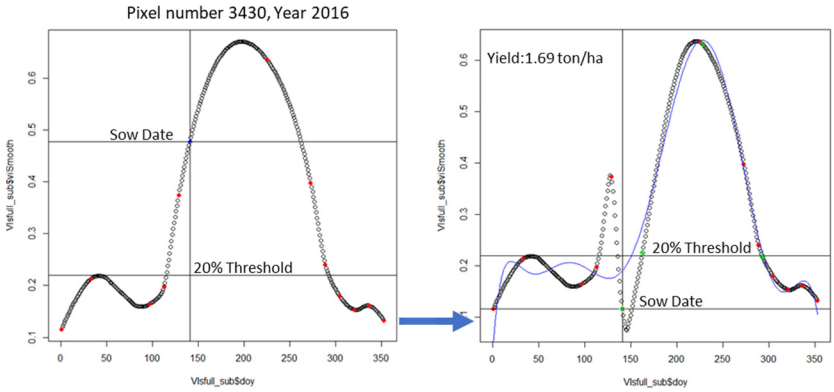


Figure S5 An example of the necessity and conditional setting to add actual sowing dates (step 2 in section 3.1.2). The left panel shows an extremely high NDVI value on actual sow date based on the smoothed daily NDVI sequence. There was only one available time point in the first half of the curve (between sowing date to peak NDVI), which led to inaccurate estimation of an extended growing season that started in early April. As a result, the integrated NDVIs in step 5 in section 3.1.1 would be over-estimated. The right panel demonstrates that adding the sowing date has constrained the shape of the fitted curve without being overly influenced by NDVI points prior to the actual farming season.

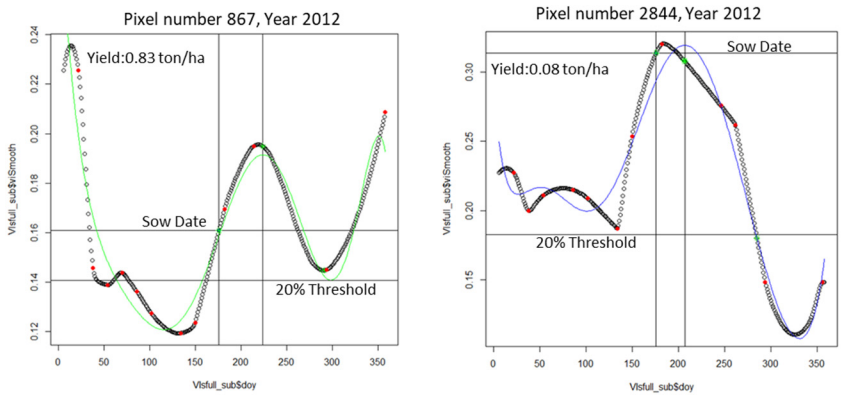


Figure S6 Two example pixels for demonstrating reasons for pixel removal. The left panel shows a pixel that meets the requirement for pixel removal at step 1 in section 3.1.2. There were only two available time points for NDVI in the 2012 growing season for that pixel. Three time points can form a triangle but not a curve. Even if the curve was fitted, it's shape was was mostly based on the NDVI observations before and after the growing season, which lacks representation of the real growing season. The right panel demonstrates a pixel that meets the requirement for pixel removal at the step 4 in section 3.1.2. For this pixel, there was only one time point available within ±11 days of the actual sowing date window. Thus, the condition for adding sow date NDVI was not activated. As a result, the reconstructed sow date NDVI is higher than the 35% threshold, which is unreasonable. The removal of such pixel is necessary to reduce bias.



Figure S7 For field J at field-average scale, the year-by-year tuning points detected by three curve fitting models: Statistical phenological detection (SPD), statistical phenology detection using sowing date (+SD), and statistical phenology detection using break of season (+BOS). The red points indicate the start of season (SOS) estimated from the three statistical models.

For the training site J, Figure S7 shows that the years that meet the conditions to include farmer-reported sow date (SD) were 2003, 2009, 2015, and 2016. The years that meet the condition to include rainfall-based break of season date (BOS) were 2003, 2006, 2015, and 2017. Among those years, only 2003, 2015, and 2016 had the estimated start of season (SOS) date changed before and after the inclusion of SD/BOS in SPD. Year 2003 has the lowest NDVI availability among the wheat planting years in field J, there were only two time points available in the first half of season after SD. The extra add-on point has delayed the detected SOS date by 44 days (April 30 to June 13) and increased the POS NDVI value by 0.033 (0.707 to 0.740), relative to being fitted by SPD only. The results indicate that, although the availability of observed NDVI sequence meets the conditions to include additional information based on SD/BOS, the difference made in phenology detection was limited but effective.

Table S1. A summary of the number of pixels processed with conditional setting in phenology detection using sowing date (+SD) and break of season (+BOS) for field J.

Conditional setting	Model	Year (Count of pixels)									
		2003	2004	2006	2007	2009	2012	2013	2015	2016	2017
Availability limits exceed	+SD/BOS	15	369	50	184	72	163	98	118	123	128
>35% threshold exceed	+SD	18	0	139	22	0	438	7	0	3	9
	+BOS	21	0	0	22	32	282	0	0	30	0
Have option to add sow date	+SD	3066	196	379	308	3009	272	299	2963	2958	204
	+BOS	3066	196	3031	308	0	376	2305	2963	256	2953

Have option to add harvest date	+SD	35	24	0	2897	0	141	164	25	215	0
Threshold value switch	+SD	0	44	509	216	0	892	45	0	109	45
	+BOS	17	5	0	208	350	157	0	0	172	0
Estimated SOS changed	+SD'	2806	142	174	63	41	1221	22	223	1551	58
	+BOS	459	65	139	34	41	9	2	222	95	54
Total	Actual yield	3066	2712	3031	2897	3009	2936	2983	2963	2958	2953
	SCYM	3045	2712	3031	2897	3009	2936	2983	2963	2958	2953

'Conditional setting' is described in Figure 3. 'Estimated SOS changed' refer to the pixels where the SOS date estimated using statistical phenology detection (SPD) is earlier than the farmer reported sowing date, but have been corrected to be later than the farmer reported sowing date in the +SD or +BOS models.

Table S2. For statistical phenology detection (SPD), statistical phenology detection using sowing date (+SD) and statistical phenology detection using break of season (+BOS), the accuracy statistics (r, AIC) of integrated NDVI metrics for yield estimation.

Model	Proxy	Index	Years									
			2003	2004	2006	2007	2009	2012	2013	2015	2016	2017
SPD	POSV	r	0.253	0.276	0.259	0.390	0.073	0.148	0.531	0.368	0.249	0.301
		AIC	3039	163	-413	-1873	1249	-666	2453	1792	944	968
	iNDVI	r	0.126	0.194	0.334	0.398	0.017	0.023	0.580	0.520	0.148	0.369
		AIC	3193	273	-560	-1895	1264	-603	2216	1290	1068	817
	GLAD	r	0.123	0.185	0.350	0.382	0.011	0.019	0.547	0.509	0.214	0.374
		AIC	3195	283	-597	-1852	1265	-602	2382	1334	995	803
	VLAD	r	0.123	0.186	0.285	0.401	0.023	0.023	0.596	0.464	0.076	0.354
		AIC	3195	282	-458	-1902	1263	-603	2131	1504	1117	853
+SD	POSV	r	0.249	0.276	0.251	0.388	0.073	0.156	0.532	0.369	0.258	0.303
		AIC	3046	162	-511	-1854	1249	-550	2441	1790	930	958
	iNDVI	r	0.124	0.182	0.323	0.399	0.018	0.071	0.576	0.539	0.164	0.372
		AIC	3194	286	-643	-1884	1264	-498	2231	1207	1053	802
	GLAD	r	0.134	0.154	0.310	0.381	0.017	0.058	0.543	0.504	0.157	0.376
		AIC	3187	312	-617	-1836	1264	-494	2389	1356	1060	792
	VLAD	r	0.116	0.206	0.326	0.403	0.018	0.081	0.592	0.541	0.167	0.360
		AIC	3201	260	-649	-1895	1264	-502	2144	1197	1050	831
+BOS	POSV	r	0.262	0.276	0.259	0.389	0.077	0.150	0.531	0.369	0.241	0.300
		AIC	3024	162	-413	-1860	1211	-670	2452	1789	930	971
	iNDVI	r	0.146	0.183	0.362	0.399	0.019	0.028	0.581	0.540	0.186	0.370
		AIC	3176	285	-628	-1888	1228	-606	2210	1200	1002	814
	GLAD	r	0.096	0.161	0.350	0.385	0.013	0.023	0.553	0.505	0.241	0.374
		AIC	3214	306	-599	-1850	1229	-605	2354	1352	930	803
	VLAD	r	0.179	0.201	0.361	0.402	0.025	0.029	0.595	0.545	0.121	0.358
		AIC	3142	266	-624	-1895	1227	-606	2135	1181	1062	845

* The red numbers indicate the proxy with minimum AIC for yield estimation within-field for each model in each year.

References:

1. Fritsch, F.N. and R.E. Carlson, *Monotone Piecewise Cubic Interpolation.* SIAM Journal on Numerical Analysis, 1980. **17**(2): p. 238-246.
2. Stineman, R.W., *A consistently well-behaved method of interpolation.* Creative Computing, 1980. **6**(7): p. 54-57.

Supplementary Material File S2: For testing field M, the comparisons of estimated and actual yields using statistical phenology detection (SPD), statistical phenology detection incorporates sowing date (+SD), statistical phenology detection incorporates break of season (+BOS) and scalable crop yield mapper (SCYM).

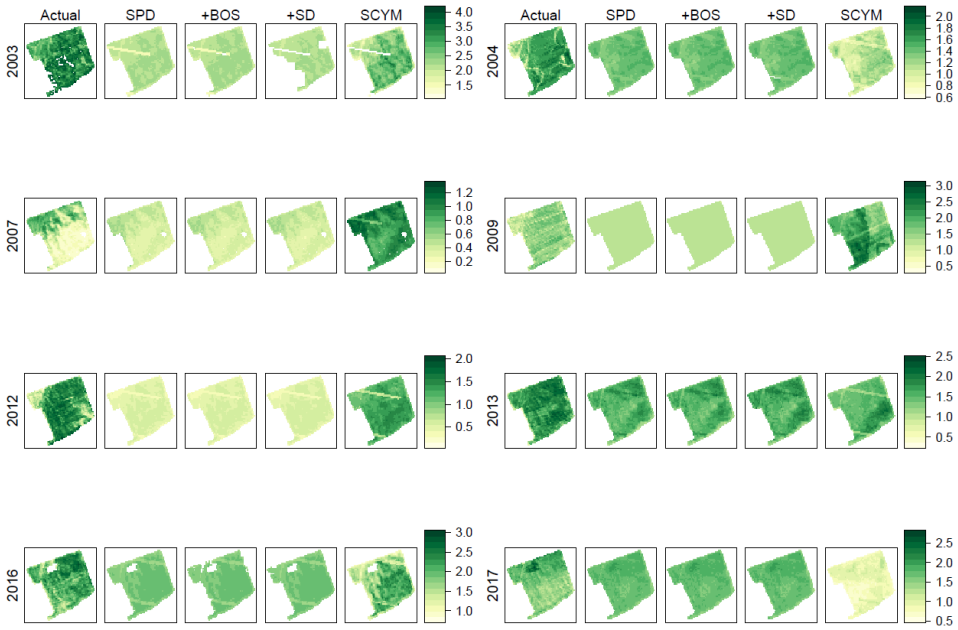


Figure S1. For field site M, actual and estimated yield maps using each of the four models: statistical phenology detection (SPD), statistical phenology detection incorporates break of season (+BOS), statistical phenology detection incorporates sowing date (+SD), and scalable crop yield mapper (SCYM).

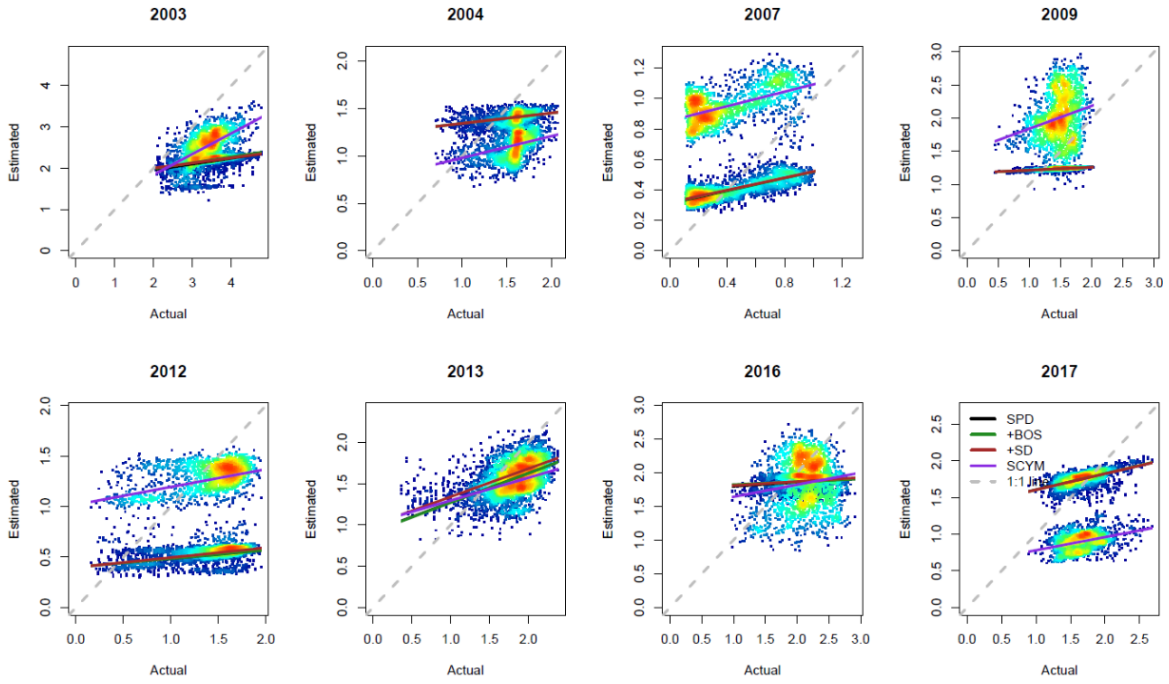


Figure S2. For field M, relationship between actual and estimated yield using each of the four models: Statistical phenology detection (SPD), statistical phenology detection incorporates sowing date (+SD), statistical phenology detection incorporates break of season (+BOS) and scalable crop yield mapper (SCYM). The colours from blue to red indicate the density of the dots from low to high.